Analysis of panels and limited dependent variable models

In honour of G.S. Maddala

Edited by
CHENG HSIAO, KAJAL LAHIRI,
LUNG-FEI LEE, AND
M. HASHEM PESARAN

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge CB2 1RP, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK http://www.cup.cam.ac.uk
40 West 20th Street, New York, NY 10011-4211, USA http://www.cup.org
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

First published 1999

Printed in the United Kingdom at the University Press, Cambridge

Typeset in Monotype Times NR 10/12 [SE]

A catalogue record for this book is available from the British Library

Library of Congress cataloguing in publication data applied for

ISBN 0 521 63169 6 hardback

Contents

List of contributors *page* vii
Foreword ix

Introduction 1

1 **A note on left censoring** 7
Takeshi Amemiya

2 **Autoregressive models with sample selectivity for panel data** 23
Manuel Arellano, Olympia Bover, and José M. Labeaga

3 **Mixture of normals probit models** 49
John Geweke and Michael Keane

4 **Estimation of dynamic limited-dependent rational expectations models** 79
Lung-Fei Lee

5 **A Monte Carlo study of EC estimation in panel data models with limited dependent variables and heterogeneity** 114
Mahmoud A. El-Gamal and David M. Grether

6 **Properties of alternative estimators of dynamic panel models: an empirical analysis of cross-country data for the study of economic growth** 136
Marc Nerlove

7 **Modified generalized instrumental variables estimation of panel data models with strictly exogenous instrumental variables** 171
Seung Chan Ahn and Peter Schmidt

8 **Expectations of expansions for estimators in a dynamic panel data model: some results for weakly exogenous regressors** 199
Jan F. Kiviet

9 **Re-examining the rational expectations hypothesis using panel data on multi-period forecasts** 226
Anthony Davies and Kajal Lahiri

10 **Prediction from the regression model with one-way error components** 255
Richard T. Baillie and Badi H. Baltagi

11 **Bayes estimation of short-run coefficients in dynamic panel data models** 268
Cheng Hsiao, M. Hashem Pesaran, and A. Kamil Tahmiscioglu

12 **Bias reduction in estimating long-run relationships from dynamic heterogeneous panels** 297
M. Hashem Pesaran and Zhongyun Zhao

CV of G. S. Maddala 323
Index 334

1 A note on left censoring

TAKESHI AMEMIYA

1 Introduction

Left censoring occurs in a duration model when a statistician observes only those spells which either are in the middle of continuation at the time of the first observation or start during the observation period. It is assumed that the statistician has no record of those spells which had ended by the time of the first observation. A special treatment of the problem is necessary because ignoring left censoring will overestimate the mean duration as longer spells tend to be observed more frequently than shorter spells. This is called selectivity bias.

Different cases of left censoring arise depending on the following considerations: (1) Spells in the middle of continuation at the time of the first observation are either completely or partially observed. Suppose such a spell started at s, continued on to 0 (the time of the first observation), and ended at t. The statistician may observe only s (by asking how long the spell had lasted), only t, or both. (2) Spells which start after the time of the first observation are either observed or not observed. (3) For a single individual we either observe a single spell or a sequence of spells in different states.

In each possible case we will consider how the selectivity bias is eliminated. If the model is fully specified, this is accomplished by the method of maximum likelihood estimation, which is fully efficient. However, in certain situations, a less efficient but more robust method, which does not require the full knowledge of the model specification, may be possible and desirable.

Although we treat the case of a homogeneous population, the adjustment for a heterogeneous population is simple as it will be indicated in appropriate places.

The problem of left censoring is dealt with only scantily in the general

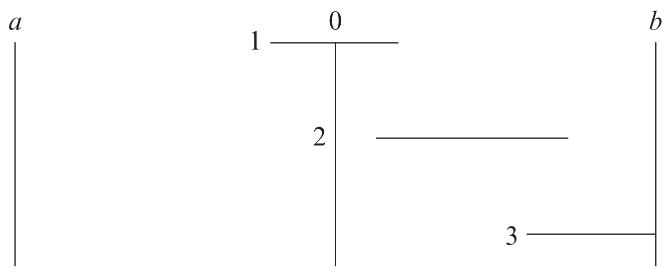
The first version of this chapter appeared as CEPR Publication No. 235 on March 1991. The author is indebted to Aaron Edlin and Tony Lancaster for helpful comments.

statistical literature. For example, standard textbooks on duration analysis such as Kalbfleisch and Prentice (1980) or Cox and Oakes (1984) devote less than a page to the problem. Miller (1981) mentions only a different kind of left censoring from what we discuss here. One can find more discussion in the econometric literature (for example, see Lancaster (1979), Flinn and Heckman (1982), Ridder (1984), and Amemiya (1985)). Here we try to give a more complete, unified treatment of the subject.

2 A single state model

The duration data are generated according to the following scheme: a duration starts in an interval $[a, b]$, which encloses 0, and the starting time X is distributed according to density $h(x)$. Duration T is distributed according to density $f(t)$ and distribution function $F(t)$. We assume that X and T are independent. The statistician observes only those spells which end or are censored after 0. We will consider three types of left censoring and for each type will derive the likelihood function assuming a homogeneous population. The result can be easily modified for the case of a heterogeneous population, as we will indicate below.

Type 1 left censor



Here the spell that was going on at time 0 is completely observed. Three kinds of spells are depicted in the above figure; we will write the likelihood function as a product of three parts corresponding to the three kinds. Each part is to be divided by the probability of observing a spell. Define

$$A_1 = \{x,t \mid t > -x,\ 0 > x > a\} \text{ and } A_2 = \{x,t \mid x > 0\}.$$

Then

$$P_1 \equiv P(A_1) = \int_a^0 h(x)[1 - F(-x)]dx \tag{1}$$

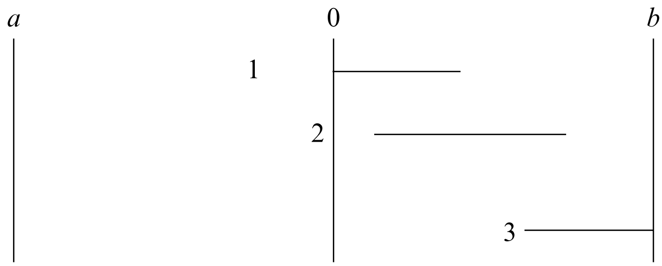
$$P_2 \equiv P(A_2) = \int_0^b h(x)dx. \tag{2}$$

The probability of observing a spell, denoted by P, is $P_1 + P_2$. Finally, the likelihood function can be written as

$$L_1 = \prod_1 h(x_i)f(t_i)\prod_2 h(x_i)f(t_i)\prod_3 h(x_i)[1-F(b-x_i)]\prod_{all} P^{-1}. \tag{3}$$

Note that the first and second kinds of spells are treated symmetrically. In the next section we will show that dividing the first part by P_1 and the second and third part by P_2 leads to a consistent but less-efficient estimator.

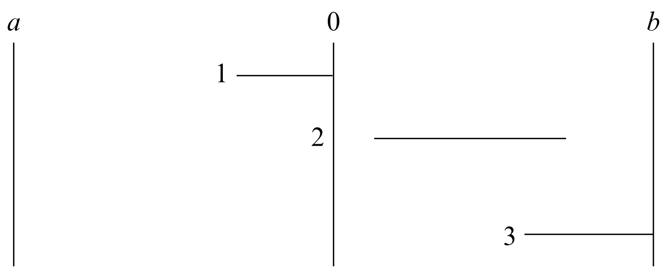
Type 2 left censor



Here the spell that was going on at time 0 is observed only after 0. The likelihood function differs from (3) only in its first part and is given by

$$L_2 = \prod_1 \int_a^0 h(x)f(t_i - x)dx \prod_2 h(x_i)f(t_i)\prod_3 h(x_i)[1-F(b-x_i)]\prod_{all} P^{-1}. \tag{4}$$

Type 3 left censor



Here the spell that was going on at time 0 is observed only up to 0. Again, the likelihood function differs from (3) and (4) only in its first part.

$$L_3 = \prod_1 h(x_i)[1 - F(t_i)] \prod_2 h(x_i) f(t_i) \prod_3 h(x_i)[1 - F(b - x_i)] \prod_{all} P^{-1}. \quad (5)$$

So far we have assumed a homogeneous population. The necessary adjustment for a heterogeneous population is straightforward. Merely add subscript i to h, f, F, a, and b, and hence also to P. Otherwise, the likelihood function (3), (4), or (5) is unchanged.

3 Why divide by P

Now we answer the question posed after equation (3): Why is it less efficient to divide the first part by P_1 and the second and third part by P_2? We will consider Type 1 left censor; the other types can be similarly analyzed. For simplicity we will assume that there is no right censoring. Therefore, there are only two kinds of spells and the spell which reaches b is observed until its end. In this case the correct likelihood function is (3) except the third part. We will first give a heuristic and then a rigorous argument.

Rewrite (3) as

$$\begin{aligned} L_1 &= \prod_1 h(x_i) f(t_i) \prod_2 h(x_i) f(t_i) \prod_{all} P^{-1} \\ &= \prod_1 h(x_i)[f(t_i)] P_1^{-1} \prod_2 h(x_i) f(t_i) P_2^{-1} \prod_1 P_1/P \prod_2 P_2/P \equiv L_{11} L_{12}, \quad (6) \end{aligned}$$

where L_{11} consists of the first two products. From the above it is clear that dividing the two parts separately by P_1 and P_2 means ignoring L_{12}. It means ignoring information that a particular spell is either the first kind or the second kind. The estimator that maximizes L_{11} is a conditional maximum likelihood estimator; therefore, it is consistent but less efficient.

To advance a rigorous argument, we must introduce a parameter vector θ to estimate. Although we will treat θ as a scalar in the subsequent analysis, an extension to the vector case is obvious. Suppose f depends on θ but h does not. Taking the natural logarithm of the first line of (6) and ignoring h because it does not depend on θ, we have

$$\log L_1 = \sum_{i=1}^{n} \log f(t_i) - n \log P \quad (7)$$

where n is the number of observed spells. Differentiating (7) with respect to θ and noting P_2 does not depend on θ, we have

$$\frac{\partial \log L_1}{\partial \theta} = \sum_{i=1}^{n} \frac{1}{f} \frac{\partial f}{\partial \theta} - \frac{n}{P} \frac{\partial P_1}{\partial \theta}. \quad (8)$$

Although it is not necessary to do so because a maximum likelihood estimator is generally consistent, we can directly check it by noting

$$E\frac{1}{N}\frac{\partial \log L_1}{\partial\theta}=\int_A \frac{1}{f}\frac{\partial f}{\partial\theta}hfdtdx-\frac{\partial P_1}{\partial\theta}=\frac{\partial}{\partial\theta}\int_A hfdtdx-\frac{\partial P_1}{\partial\theta}=0 \tag{9}$$

where $A=A_1\cup A_2$ and N is the total number of spells both observed and unobserved.

Differentiating (8) again with respect to θ

$$\frac{\partial^2 \log L_1}{\partial\theta^2}=-\sum_{i=1}^{n}\frac{1}{f^2}\left[\frac{\partial f}{\partial\theta}\right]^2+\sum_{i=1}^{n}\frac{1}{f}\frac{\partial^2 f}{\partial\theta^2}+\frac{n}{P^2}\left[\frac{\partial P_1}{\partial\theta}\right]^2-\frac{n}{P}\frac{\partial^2 P_1}{\partial\theta^2}. \tag{10}$$

Therefore

$$-E\frac{1}{N}\frac{\partial^2 \log L_1}{\partial\theta^2}=\int_A \frac{1}{f}\left[\frac{\partial f}{\partial\theta}\right]^2 hdtdx-\frac{1}{P}\left[\frac{\partial P_1}{\partial\theta}\right]^2. \tag{11}$$

As usual, the asymptotic variance of $\sqrt{N}(\hat{\theta}-\theta)$ is given by the inverse of (11).

Next, taking the log of L_{11} and ignoring the terms that do not depend on θ

$$\log L_{11}=\sum_{i=1}^{n}\log f(t_i)-n_1\log P_1-n_2\log P_2. \tag{12}$$

Therefore

$$\frac{\partial \log L_{11}}{\partial\theta}=\sum_{i=1}^{n}\frac{1}{f}\frac{\partial f}{\partial\theta}-\frac{n_1}{P_1}\frac{\partial P_1}{\partial\theta}. \tag{13}$$

As in (9), we can show

$$E\frac{1}{N}\frac{\partial \log L_{11}}{\partial\theta}=0 \tag{14}$$

which implies the consistency of the estimator that maximizes L_1. As in the case of L, the asymptotic variance is the inverse of

$$-E\frac{1}{N}\frac{\partial^2 \log L_{11}}{\partial\theta^2}=\int_A \frac{1}{f}\left[\frac{\partial f}{\partial\theta}\right]^2 hdtdx-\frac{1}{P_1}\left[\frac{\partial P_1}{\partial\theta}\right]^2. \tag{15}$$

Since L_{11} is a conditional likelihood function, this result is correct, but one can also directly show the equality of (15) to $N^{-1}V(\partial\log L_{11}/\partial\theta)$.

By comparing (15) with (11), one can readily see the inefficiency of the estimator that maximizes L_{11} relative to the true maximum likelihood estimator.

4 A simple example

In this section we will evaluate the variances of the above two maximum likelihood estimators in a very simple duration model. We will also calculate the degree of inconsistency of an estimator which does not correct for selective bias. We assume that spells start either at -1 or at 0 with equal probabilities. From either starting point, a spell lasts for 1.5 with probability p and 0.5 with probability $1-p$. We observe only those spells which end after 0. Supposing we observe m_1 spells which started at -1 and lasted for 1.5, n_1 spells which started at 0 and lasted for 1.5 and n_2 spells which started at 0 and lasted for 0.5, how should we estimate p?

Ignoring selectivity

This estimator maximizes the product of unadjusted probabilities

$$S = p^{m_1+n_1}(1-p)^{n_2}. \tag{16}$$

Therefore, $\tilde{p} = (m_1+n_1)/(m_1+n_1+n_2)$ and plim $\tilde{p} = p - (p^2-p)/(1+p)$. The parameter p is overestimated because the short spells that started at -1 were not observed.

Maximizing conditional LF

The conditional likelihood function (aside from a constant term), the conditional maximum likelihood estimator, and its asymptotic distribution are given by

$$L_1 = p^{n_1}(1-p)^{n_2} \tag{17}$$

$$\hat{p}_1 = \frac{n_1}{n_1+n_2} \tag{18}$$

$$\sqrt{N}(\hat{p}_1 - p) \to N\{0, 2p(1-p)\} \tag{19}$$

where N is the total number of spells including those which are not observed. Note that since (18) defines the estimator explicitly, (18) can be verified directly by a central limit theorem. However, the asymptotic variance may also be obtained as the inverse of the information matrix as we did in the previous section. The same remark applies to the next estimator.

Maximizing full LF

The full likelihood function (aside from a constant term), the maximum likelihood estimator, and its asymptotic distribution are given by

$$L = p^{m_1+n_1}(1-p)^{n_2}(1+p)^{-(m_1+n_1+n_2)} \tag{20}$$

$$\hat{p} = \frac{m_1+n_1}{m_1+n_1+2n_2} \tag{21}$$

$$\sqrt{N}(\hat{p}-p) \rightarrow N\{0, p(1-p)(1+p)\}. \tag{22}$$

From (19) and (22), we see the extent of the inefficiency of the conditional maximum likelihood estimator.

5 Observe only spells continuing at 0

In this section we consider the model in which duration data are generated by the same scheme as in section 2 but we observe only those spells which are continuing at time 0. Since we can concentrate on the starting time in the interval $[a, 0]$, we will assume that the support of density $h(x)$ is $[a, 0]$. We assume a homogeneous population, but the adjustment for the case of a heterogeneous population is simple as in section 2. As before, we consider three types of censoring. For each type the likelihood function is the first part of the likelihood function of section 2 divided this time not by P but by P_1. Thus

$$L_1 = \prod h(x_i) f(t_i) P_1^{-1} \tag{23}$$

$$L_2 = \prod \int_a^0 h(x) f(t_i - x) dx \; P_1^{-1} \tag{24}$$

$$L_3 = \prod h(x_i)[1 - F(t_i)] P_1^{-1}. \tag{25}$$

When we take $h(x)$ as uniform density and take the limit of a going to $-\infty$, we obtain the three formulae (11.2.75), (11.2.76), and (11.2.71), respectively, given in Amemiya (1985, p. 448).

6 Method which does not require starting-time distribution

Next, we will consider Type 1 left censor as adapted in the model of section 5 above, for which the full likelihood function is (23), and discuss a consistent estimator which does not require the knowledge of $h(x)$. It is the estimator that maximizes the product of conditional densities of t given x. This estimator was used by Lancaster (1979), and its properties were studied by Ridder (1984). Such an estimator is useful because it is often difficult for a researcher to specify $h(x)$ correctly. The aforementioned formulae in Amemiya do not depend on $h(x)$ but this was accomplished by an arbitrary assumption of its uniformity. Although we treat the case of a homogeneous population, the adjustment for a heterogeneous population is simple as in section 2. This estimator maximizes

$$L^* = \prod_{i=1}^{n} \frac{f(t_i)}{1 - F(-x_i)}. \tag{26}$$

Clearly, it is a conditional maximum likelihood estimator. Taking the logarithm

$$\log L^* = \sum_{i=1}^{n} \log f(t_i) - \sum_{i=1}^{n} \log[1 - F(-x_i)]. \tag{27}$$

Differentiating (27) with respect to θ

$$\frac{\partial \log L^*}{\partial \theta} = \sum_{i=1}^{n} \frac{1}{f}\frac{\partial f}{\partial \theta} + \sum_{i=1}^{n} \frac{1}{1-F}\frac{\partial F}{\partial \theta}. \tag{28}$$

Consistency follows from noting

$$\begin{aligned} E\frac{1}{N}\frac{\partial \log L^*}{\partial \theta} &= \int_{A_1} \frac{1}{f}\frac{\partial f}{\partial \theta} hf dt dx + \int_{A_1} \frac{1}{1-F}\frac{\partial f}{\partial \theta} hf dt dx \\ &= \frac{\partial}{\partial \theta}\int_a^0 h(x)[1 - F(-\mathrm{x})]dx + \frac{\partial}{\partial \theta}\int_a^0 h(x)F(-\mathrm{x})dx \\ &= \frac{\partial}{\partial \theta}\int_a^0 h(x)dx \qquad (29) \\ &= 0. \end{aligned}$$

Differentiating (28) again with respect to θ

$$\begin{aligned} \frac{\partial^2 \log L^*}{\partial \theta^2} = &-\sum_{i=1}^{n} \frac{1}{f^2}\left[\frac{\partial f}{\partial \theta}\right]^2 + \sum_{i=1}^{n} \frac{1}{f}\frac{\partial^2 f}{\partial \theta^2} + \sum_{i=1}^{n} \frac{1}{(1-F)^2}\left[\frac{\partial F}{\partial \theta}\right]^2 \\ &+ \sum_{i=1}^{n} \frac{1}{1-F}\frac{\partial^2 F}{\partial \theta^2}. \end{aligned} \tag{30}$$

Using

$$E\frac{1}{f^2}\left[\frac{\partial f}{\partial \theta}\right]^2 = \int_{A_1} \frac{1}{f}\left[\frac{\partial f}{\partial \theta}\right]^2 hdtdx \tag{31}$$

$$E\frac{1}{f}\frac{\partial^2 f}{\partial \theta^2} = \frac{\partial^2}{\partial \theta^2}\int_a^0 h(x)[1 - F(-x)]dx \tag{32}$$

$$E\frac{1}{(1-F)^2}\left[\frac{\partial F}{\partial \theta}\right]^2=\int_a^0 \frac{1}{1-F}\left[\frac{\partial F}{\partial \theta}\right]^2 hdtdx \tag{33}$$

$$E\frac{1}{1-F}\frac{\partial^2 F}{\partial \theta^2}=\frac{\partial^2}{\partial \theta^2}\int_a^0 h(x)[F(-x)]dx \tag{34}$$

we have

$$-E\frac{1}{N}\frac{\partial^2 \log L^*}{\partial \theta^2}=\int_{A_1}\frac{1}{f}\left[\frac{\partial f}{\partial \theta}\right]^2 hdtdx-\int_a^0 \frac{1}{1-F}\left[\frac{\partial F}{\partial \theta}\right]^2 hdtdx. \tag{35}$$

The information matrix of the full maximum likelihood estimator that maximizes (23) is analogous to (15) except that the integration in the right-hand side is over the set A_1. Thus

$$-E\frac{1}{N}\frac{\partial^2 \log L_1}{\partial \theta^2}=\int_{A_1}\frac{1}{f}\left[\frac{\partial f}{\partial \theta}\right]^2 hdtdx-\frac{1}{P_1}\left[\frac{\partial P_1}{\partial \theta}\right]^2. \tag{36}$$

That (36) is greater than (35) follows from the Cauchy–Schwartz inequality

$$E_xU^2\,E_xV^2 \geq (E_xUV)^2 \tag{37}$$

where $E_x(Z)=\int_a^0 zh(x)dx$, $U=\frac{1}{\sqrt{1-F}}\frac{\partial F}{\partial \theta}$, and $V=\sqrt{1-F}$.

Unfortunately, a simple consistent estimator which does not require the knowledge of $h(x)$ has not been found for Type 2 or Type 3 left censor in the model of section 5. In the model of section 2, there exists an even simpler consistent estimator of θ which does not require the knowledge of h, provided h does not depend on θ. It works for any of the three types and simply amounts to maximizing the second and third product each divided by P_2: namely

$$\text{Maximize} \prod_2 h(x_i)f(t_i)P_2^{-1}\prod_3 h(x_i)[1-F(b-x_i)]P_2^{-1}.$$

In the case of Type 1 left censor, this method can be combined with the method proposed earlier to obtain a more efficient estimator: namely, that which maximizes

$$L^*=\prod_1 h(x_i)f(t_i)[1-F(-x_i)]^{-1}\prod_2 h(x_i)f(t_i)P_2^{-1}\prod_3 h(x_i) \\ [1-F(b-x_i)]P_2^{-1}. \tag{38}$$

If $h(x)$ does not depend on θ as before, $h(x_i)$ and P_2 can be eliminated from the right-hand side of (38). The estimator is consistent and its asymptotic variance can be obtained by the same method described above.

There is an alternative method, which, although not consistent, may work reasonably well for Type 2 left censor in the model of section 5. In this method, the term after the product symbol in the right-hand side of (24) is specified to be an appropriate function of t and a new set of parameters. This method can be easily adjusted for heterogeneous population and has been used in empirical applications (see, for example, Gritz (1993)).

7 Semiparametric estimation of $h(x)$ and θ

Goto (1993 and 1996) considered the semiparametric maximum likelihood estimation of density $h(x)$ and the parameter θ that appears inside f in the model (23). He proved that when the maximizing value of $h(x)$ is inserted back into (23), one obtains the conditional likelihood function (26). Thus, we can interpret the conditional maximum likelihood estimator that maximizes (26) as the semiparametric maximum likelihood estimator of the model (23).

We will give a sketch of the proof. We are to maximize

$$L = \prod_{i=1}^{n} \frac{h(x_i) f(t_i)}{\int_a^0 h(x)[1 - F(-x)]dx} \tag{39}$$

with respect to function $h(\cdot)$. We can take $h(\cdot)$ to be a step function taking the value of h_i over the interval of length d around the observed x_i and zero elsewhere, where $d = (\sum h_k)^{-1}$. Then, since

$$\int_a^0 h(x)[1 - F(-x)]dx \cong \sum_k dh_k[1 - F(-x_k)] \tag{40}$$

we can write (39) as

$$L = \prod_{i=1}^{n} \frac{h_i f(t_i)}{\sum_{k=1}^{n} h_k (\sum_j h_j)^{-1}[1 - F(-x_k)]} \tag{41}$$

Therefore, we should maximize

$$\prod_{i=1}^{n} \frac{h_i f(t_i)}{\sum_k h_k[1 - F(-x_k)]} \text{ subject to } \sum_k h_k = 1. \tag{42}$$

Differentiating the logarithm of the above with respect to each h_i and setting the derivative equal to zero, we obtain

$$\frac{1}{h_i} - \frac{na_i}{\sum h_k a_k} = 0 \tag{43}$$

where we have defined $a_i = 1 - F(-x_i)$. These can be rewritten as

$$h_i = \left(\frac{a_1}{a_i}\right) h_1,\ i = 1, 2, \ldots, n. \tag{44}$$

Summing both sides of the above over i and noting $\sum_k h_k = 1$, we obtain

$$h_i = \frac{1}{a_i \sum_k (a_k)^{-1}}. \tag{45}$$

Finally, putting (45) back into the maximand in (42) yields

$$\frac{1}{n^n} \prod_{i=1}^{n} \frac{f(t_i)}{1 - F(-x_i)} \tag{46}$$

which is the same as (26) except for a factor which does not depend on θ.

8 Separate estimation of $h(x)$

The density $h(x)$ may also be estimated using a sample either independent of that used to estimate θ or an augmented sample including that used to estimate θ. Nickell (1979) used the method but did not analyze its effect on the asymptotic distribution of the maximum likelihood estimator of θ.

We consider the estimator of θ that maximizes

$$\hat{L}_1 = \prod f(t_i) \hat{P}_1^{-1} \tag{47}$$

where $\hat{P}_1 = \int_a^0 \hat{h}(x)[1 - F(-x)]dx$. As $\hat{h}(x)$, we can, for example, use the following simple method: divide the interval $[a, 0]$ into small intervals of length d and for each interval estimate $h(x)$ by the relative frequency of the workers who became unemployed in the augmented sample. If the size of the independent or augmented sample is K, we must have $d \to 0$ and $dK \to \infty$. In this case we have (see Bickel and Doksum (1977, p. 385))

$$|\hat{h}(x) - h(x)| = O\left(\frac{1}{\sqrt{dK}}\right) \tag{48}$$

provided that $h(x)$ is continuous in $[a, 0]$.

If we denote this estimator by $\hat{\theta}$, its asymptotic distribution can be obtained from

$$\sqrt{N}(\hat{\theta}-\theta)=-\left(\frac{1}{\sqrt{N}}\frac{\partial\log\hat{L}_1}{\partial\theta}\right)\left(\frac{1}{N}\frac{\partial^2\log\hat{L}_1}{\partial\theta^2}\right)^{-1}. \tag{49}$$

The second-derivative term above divided by N will converge to the same limit as if h were not estimated, provided that $d\to 0$ and $dK\to\infty$. So, here, we will only consider the first derivative part. We have

$$\frac{1}{\sqrt{N}}\frac{\partial\log\hat{L}_1}{\partial\theta}\cong\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(D\frac{1}{f}\frac{\partial f}{\partial\theta}-\frac{\partial P_1}{\partial\theta}\right)-\frac{1}{P_1}\frac{\partial P_1}{\partial\theta}\frac{1}{\sqrt{N}}\sum_{1}^{N}(D-P_1)$$
$$+\int_a^0\sqrt{N}\,[\hat{h}(x)-h(x)]\Psi(x)dx, \tag{50}$$

where $\Psi(x)=\partial F(-x)/\partial\theta+P_1^{-1}(\partial P_1/\partial\theta)[1-F(-x)]$ and $D_i=1$ if $T_i>-X_i>0$ and 0 otherwise. From (48) and (50) it is clear that only when K goes to infinity at the rate faster than N, estimating h has no effect on the asymptotic distribution.

It is better to estimate P_1 by $\hat{P}_1=\int_a^0[1-F(-x)]d\hat{H}(x)$, where $\hat{H}$ is the empirical distribution function. Thus, $\hat{P}_1=\frac{1}{K}\sum_{i=1}^{K}[1-F(-x_i)]$. In this case, we can show

$$\frac{1}{\sqrt{N}}\frac{\partial\log\hat{L}_1}{\partial\theta}\cong\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(D\frac{1}{f}\frac{\partial f}{\partial\theta}-\frac{\partial P_1}{\partial\theta}\right)+\sqrt{N}\int_a^0\frac{\partial F}{\partial\theta}d[\hat{H}(x)-H(x)] \tag{51}$$

but since

$$\sqrt{K}\int_a^0\frac{\partial F}{\partial\theta}d[\hat{H}(x)-H(x)]\to N[0,\ V(\partial F/\partial\theta)], \tag{52}$$

the asymptotic variance of this estimator adds $cV(\partial F/\partial\theta)$ to that of the maximum likelihood estimator with the known $h(x)$, where $cK=N$ for some constant c (see Prakasa Rao (1987, p. 391)).

9 Two-states model

In this section we consider a model in which an individual starts a spell in either state 1 or state 2 at time a, with probability p and $1-p$ respectively. The transition from state 1 to 2 occurs with density function $f_1(t)$ and distribution function $F_1(t)$. The transition from 2 to 1 is done according to $f_2(t)$

and $F_2(t)$. The statistician starts observing spells at time 0. As in section 2, the spells that are continuing at time 0 will be completely or partially observed, leading to the three types of censoring. The subsequent spells of an individual, after the spell that is continuing at 0 terminates, will be observed up to time b. However, unlike in section 2, possible selectivity bias need not be taken into account for these subsequent spells, because the starting time of a subsequent spell is determined by the end of the preceding spell. For this reason this model has more similarity to the model of section 5. Although we treat the case of a homogeneous population, the adjustment for a heterogeneous population is simple as in section 2.

We need to distinguish the following two cases: (1) the statistician observes spells in both states at time 0; (2) the statistician observes only spells in one of the states at time 0. We will assume that the subsequent spells of both states will be observed, but this is not a crucial assumption.

Observe both states

The likelihood functions corresponding to the three types of censoring can be written as follows:

$$L_1 = \prod_1 h_1(x_i) f_1(t_i) \prod_2 h_2(x_i) f_2(t_i) \tag{53}$$

$$L_2 = \prod_1 \int_a^0 h_1(x) f_1(t_i - x) dx \prod_2 \int_a^0 h_2(x) f_2(t_i - x) dx \tag{54}$$

$$L_3 = \prod_1 h_1(x_i)[1 - F_1(t_i)] \prod_2 h_2(x_i)[1 - F_2(t_i)] \tag{55}$$

where the numbers at the bottom of the product sign refer to the two states rather than two kinds of spells as in section 2.

A feature of this model which makes it more complex than the single-state model is that $h(x)$ is a very complicated function of p, f_1, and f_2. Thus, if θ is the vector of parameters that characterize f_1 and f_2, we can no longer assume that $h(x)$ does not depend on θ. This fact makes the full maximum likelihood estimator generally very complicated and increases the advisability of a simple consistent estimator which does not require $h(x)$. We will briefly indicate how $h_1(x)$ can be determined. It can be calculated by summing the densities of all the possible event histories prior to x. First of all, $x = a$ with probability p. Therefore, strictly speaking, we should allow for the possibility that hs that appear in the right-hand side of (53) and (55) are probabilities and the integral in (54) is a Stieltjes integral. The second possibility is that an individual starts in state 2 at time a and moves once to state 1 at time x, for which the density is

$$(1-p)f_2(x-a) \tag{56}$$

The third possibility is that an individual starts in state 1 at time a, moves to state 2, comes back to state 1 at time x, for which the density is

$$p\int_a^x f_1(y-a)f_2(x-y)dy. \tag{57}$$

Summing all these densities *ad infinitum*, $h_1(x)$ is evaluated. $h_2(x)$ can be analogously derived.

To verify the correctness of our likelihood functions, it is useful to check the consistency of the maximum likelihood estimator. We will do this for (53). We have

$$\frac{\partial \log L_1}{\partial \theta} = \sum_1 \frac{1}{h_1}\frac{\partial h_1}{\partial \theta} + \sum_1 \frac{1}{f_1}\frac{\partial f_1}{\partial \theta} + \sum_2 \frac{1}{h_2}\frac{\partial h_2}{\partial \theta} + \sum_2 \frac{1}{f_2}\frac{\partial f_2}{\partial \theta} \tag{58}$$

$$E\frac{1}{N}\frac{\partial \log L_1}{\partial \theta} = \frac{\partial P_1}{\partial \theta} + \frac{\partial P_2}{\partial \theta} = 0 \tag{59}$$

where P_1 is the probability an individual is in state 1 at time 0 and is given by

$$P_1 = \int_a^0 h_1(x)[1-F_1(-x)]dx \tag{60}$$

Clearly, $P_2 = 1 - P_1$.

Observe one state

Without loss of generality, we assume we observe only spells in state 1 at time 0. The likelihood functions corresponding to the three types of censoring can be written as follows

$$\underline{L}_1 = \prod h_1(x_i)f_1(t_i)P_1^{-1} \tag{61}$$

$$\underline{L}_2 = \prod \int_a^0 h_1(x)f_1(t_i-x)dxP_1^{-1} \tag{62}$$

$$\underline{L}_3 = \prod h_1(x_i)[1-F_1(t_i)]P_1^{-1}. \tag{63}$$

These likelihood functions are not any simpler than those in the previous case.

Method which does not require starting-time distribution

As in section 6, this method works only for type 1 left censor. This estimator maximizes the following conditional likelihood functions: L_1^* for the case of observing both states and $\underline{L}_1^*$ for the case of observing only state 1.

$$L_1^* = \prod_1 \frac{f_1(t_i)}{1 - F_1(-x_i)} \prod_2 \frac{f_2(t_i)}{1 - F_2(-x_i)} \tag{64}$$

$$\underline{L}_1^* = \prod \frac{f_1(t_i)}{1 - F_1(-x_i)}. \tag{65}$$

The consistency of these estimators can be shown by tracing the argument in section 6.

It is instructive to recognize the term after the product symbol in (61) and (65) as successive conditional likelihood functions of the term after the first product symbol in (53). The term after the first product symbol in (53) can be written as the product of three terms as follows

$$hf = \frac{f}{1-F} \frac{h(1-F)}{P} P \tag{66}$$

where we have omitted the subscript 1 from all the letters. The product of the first two terms in the right-hand side of (66) gives the term in (61) and the first term alone in the right-hand side of (66) gives the term in (65).

References

Amemiya, T. (1985), *Advanced Econometrics*, Cambridge, Mass.: Harvard University Press.

Bickel, P. and K. Doksum (1977), *Mathematical Statistics*, Oakland: Holden-Day.

Cox, D.R. and D. Oakes (1984), *Analysis of Survival Data*, London: Chapman and Hall.

Flinn, C.J. and J.J. Heckman (1982), "Models for the Analysis of Labor Force Dynamics," *Advances in Econometrics*, 1: 35–95.

Goto, F. (1993), "Consistency and Efficiency of Semiparametric Estimators in Left Censored Duration Models," Unpublished Ph.D. Thesis. Stanford University.

(1996), "Achieving Semiparametric Efficiency Bounds in Left Censored Duration Models," *Econometrica*, 64: 439–442.

Gritz, R.M. (1993), "The Impact of Training on the Frequency and Duration of Employment," *Journal of Econometrics*, 57: 21–51.

Kalbfleisch, J.D. and R.L. Prentice (1980), *The Statistical Analysis of Failure Time Data*, New York: John Wiley & Sons.

Lancaster, T. (1979), "Econometric Methods for the Duration of Unemployment," *Econometrica*, 47: 939–956.

Miller, R.G. (1981), *Survival Analysis*, New York: John Wiley & Sons.

Nickell, S. (1979), "Estimating the Probability of Leaving Unemployment," *Econometrica*, 47: 1249–1266.

Prakasa Rao, B.L.S. (1987), *Asymptotic Theory of Statistical Inference*, New York: John Wiley & Sons.

Ridder, G. (1984), "The Distribution of Single-Spell Duration Data," in G.R. Neumann and N.C. Wertergaard-Nielsen (eds.), *Studies in Labor Market Dynamics*, Springer Verlag.